

評判分析システムのための教師あり学習を用いた意見抽出

著者	田中 俊行 , グエン ミン テイ, 中川 博之, 田原 康之, 大須賀 昭彦
雑誌名	電子情報通信学会論文誌. D, 情報・システム
巻	J94-D
号	11
ページ	1751-1761
発行年	2011-11-01
URL	http://id.nii.ac.jp/1438/00009107/

評判分析システムのための教師あり学習を用いた意見抽出

田中 俊行^{†a)} グエン ミンティ[†] 中川 博之[†] 田原 康之[†]
大須賀昭彦[†]

Opinion Extraction Using Supervised Learning for Reputation Analysis System

Toshiyuki TANAKA^{†a)}, The-Minh NGUYEN[†], Hiroyuki NAKAGAWA[†],
Yasuyuki TAHARA[†], and Akihiko OHSUGA[†]

あらまし インターネットの普及に伴い、Web 上には商品やサービス（対象物）に対する多くの評判情報が蓄積されている。しかし、誰でも発信できるが故に、情報は膨大となり、それら全てに目を通すことは利用者にとって多大な負担となる。そのような背景から、レビューから意見を自動的に抽出する研究が盛んに行われており、意見をく対象物、評価視点（属性）、評価値の三つ組と捉え抽出する研究も行われている。しかしながら多くの研究は、評価視点や評価値の抽出に辞書を用いており、ジャンルごとに必要となる辞書の構築のためのコストは小さいとはいえない。また、単に辞書を用いてマッチングを行っただけでは、精度が上がらないのが現状である。そこで本論文では、教師あり学習を用いて、レビューサイトから意見を抽出する手法を提案する。提案手法は、従来の手法のように大規模な辞書をあらかじめ用意する必要がないため、コストを大幅に抑えることが可能である。実験の結果、辞書をあらかじめ用意しない既存手法と比較して、最大で適合率は約 26%、再現率は約 47%向上した。また、既存研究では個別の辞書を必要とするような他ジャンルに対して本手法を適用した結果、ほぼ変わらない精度で抽出することができ、他ジャンルへの適用の可能性を示すことができた。

キーワード 意見抽出、評判情報、機械学習

1. ま え が き

近年、インターネットの爆発的な普及により、誰でも気軽に情報を発信できる環境が整ってきた。個人が発信する主な情報として、商品・サービス（以下、対象物と呼ぶ）に対する意見が挙げられる。中でも「価格. com」や「Amazon.co.jp」などの対象物ごとにレビューのスペースが設けられているサイトにおいては数多くのレビューが投稿されている。このようなレビューは、対象物の購入や利用を考えている消費者やその対象物にかかわる企業の参考になる場合がある。例えば消費者は、対象物の良い部分・悪い部分を知ることができ、企業においては自社が提供する対象物に対する消費者の意見を知り、対象物の改善に役立てることが可能である。しかし投稿されるレビューが膨大

となってしまう場合があり、「価格. com」を例に挙げると、一つの商品に対して数百にも及ぶレビューが投稿されている。投稿されるレビューが膨大である場合、必要な情報を入手するためには、多くの文を読まなければならない、利用する消費者や企業にとっては多大な負担となる。

このような背景から、評判を抽出し、分析、要約するシステムやサービスが存在する。これらは、Web 上の CGM から意見を抽出し、対象物全体や、対象物の側面を表す『評価視点（属性）』に対する肯定意見・否定意見をまとめ、可視化するものである [3], [4]。これらのシステムやサービスの基礎となっている意見抽出の研究では、あらかじめ構築した辞書を用いることが一般的である [14]。しかし、小林ら [12] が述べているように、対象物のジャンル（ドメイン）によって評価視点、評価値となる表現が異なるため、ジャンルごとに辞書が必要であることが認識されているが、無数に存在するジャンル全てに対する辞書を構築することは、コストの観点から現実的ではない。また、大規模な辞書を用いて意見抽出を行ったとしても、評価視点と評

[†] 電気通信大学大学院情報システム学研究科，調布市
Graduate School of Information Systems, University of
Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi,
182-8585 Japan

a) E-mail: t_tanaka@ohsuga.is.uec.ac.jp

価値のペアを同定する問題もあり、あまり高い精度で抽出できていないのが現状である。

そこで本研究では、辞書を用いずに、教師あり学習を用いて意見抽出を行う手法を提案する。本論文での抽出対象は、評価視点と評価値である。評価視点とは、対象物の側面を表すものであり、評価値とは、対象物または対象物の側面に対する評価である。そして、抽出した評価視点と評価値の出現パターンに応じて評価視点-評価値の対を同定する。また、レビューの構造を利用して、ジャンルの代表的な評価視点を更に割り当て、評判分析システムに応用するための意見の構造化を図る。既存手法と比べて、本手法には以下のような利点がある。

1. 辞書を構築する必要がない。
2. 汎用評価値辞書と、我々が構築した評価視点辞書を用いた手法よりも良い精度で意見の抽出ができる。評価実験の結果、最大で適合率は約 23%、再現率は約 40%の向上が見られた。
3. 既存研究では個別に辞書を必要とするような他ジャンルに対して本手法を適用しても、同一ジャンルに適用した場合とほぼ同程度の精度で意見の抽出ができる。

以下、2. で意見の定義や本論文の目的、意見抽出の難しさについて述べ、3. では本論文の提案手法を述べる。そして 4. で評価実験、考察を行い、5. で関連研究について述べる。最後に 6. で本論文のまとめと、今後の課題について述べる。

2. 意見抽出

2.1 意見の定義

消費者によるレビューでは、「このカメラは良い」のような対象物に対しての「良い」、「悪い」だけではなく、「このカメラの起動は早い」や「このカメラのバッテリー寿命は短い」といったように評価視点を伴って具体的に記述されることがある。そこで本研究では、意見は〈対象物、評価視点（属性）、評価値〉の三つ組からなるとする。例を図 1 に示す。対象物は商品名やサービス名に該当する。評価視点は対象物の特徴や性質を表す表現であり、例えばデジタルカメラの分野では、デザイン、画質、バッテリーなどが該当する。評価値は肯定または否定の評価を表す表現であり、例えば「画質が良い」の「良い」や、「起動が早い」の「早い」が該当する。

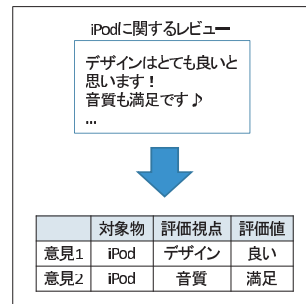


図 1 意見の例

Fig. 1 An example of opinion extraction.

2.2 本論文の目的

本論文では、意見を可視化する評判分析システムに用いるための意見を、レビューサイトから抽出することを目的とする。ただし、近年は商品ごとのレビューページが用意されるなど、対象物が一意に決まる可能性が高いため、〈評価視点、評価値〉の要素からなる組の抽出に留める。Web 上に存在する膨大なレビューからこのような組を自動的に抽出し可視化する評判分析システムを用いれば、負担を負うことなく、消費者は自らの嗜好に合わせた対象物を選択でき、企業は自社の対象物に対する消費者の詳細な意見を知り、改善に役立てることが可能である。

本論文では以下の三つの課題を設定する。

2.2.1 評価視点、評価値の抽出

第一の課題は、レビューの文中から評価視点、評価値を抽出することである。例えば、「派手ですが、質感は良いです。」という文からは、評価視点として「質感」、評価値として「派手」と「良い」が抽出されなければならない。

2.2.2 評価視点-評価値対の同定

第二の課題は、抽出した評価視点と評価値を意見とするために、評価視点-評価値の対を同定することである。先ほどの例の場合、評価視点「質感」と、評価値「良い」が対として同定されなければならない。評価値「派手」に関しては、同一文中には対となる評価視点がないと判定する必要がある。

2.2.3 抽出した意見のグルーピング

Web 上には多くのレビューが存在し、抽出される意見は膨大な数になる。評判分析システムを構築するためには、何らかの基準で意見をまとめることが必要であると考えられ、抽出した意見のグルーピングを第三の課題とする。本論文では、意見を商品ジャンルの代

表的な評価視点をを用いてまとめることを考える。そのために、評価視点-評価値対に対して、更に代表的な評価視点を割り当てる必要がある。区別するために、この代表的な評価視点を『上位評価視点』とし、文内に出現する上位評価視点以外の評価視点を『下位評価視点』とする。

2.3 意見抽出の難しさ

意見抽出には以下のような難しさがある。

1. 評価視点、評価値の表現の多様性

小林ら [9] によると、評価視点に関してはジャンルに依存する表現がほとんどであることが分かっている。評価値は対象物のジャンル（ドメイン）に強く依存しない表現も見られるが、やはり固有の表現が数多く存在する。価格. com においても数百という商品ジャンルが存在していることを考えると、全てのジャンルに対する辞書を構築することは多大なコストとなる。

2. 評価視点-評価値対の位置関係の多様性

CaboCha [6] 等を用いて構文解析を行い、係り受けを見ることによってある程度の評価視点-評価値対は同定できるが、評価視点と評価値が係り受け関係にあるのは 40% 程度であるということが Kobayashi らの研究 [5] で明らかになっている。

3. 提案手法

本論文が提案する意見抽出の手法について説明する。意見抽出は大きく三つのタスクに分けられる。

- 上位評価視点の同定
- 下位評価視点、評価値の抽出
- 下位評価視点-評価値対の同定

以下、それぞれのタスクについて説明する。

3.1 上位評価視点の同定

上位評価視点は、「価格. com」で指定されている項目を用いる。これらはジャンルごとに異なり、例えばジャンルが「デジタルカメラ」であれば、「デザイン」、「画質」、「操作性」、「バッテリー」、「携帯性」、「機能性」、「液晶」、「ホールド感」の 8 項目が設定されている。その影響もあり、レビューはそれらの項目に従い、特有のフォーマットでレビューを書く傾向にある。例を図 2 に示す。この例では「■デザイン」と書いた後の 3 行はデザインについての記述である。同様に「画質」と「操作性」について記述されている。文章を見ると、「デザイン」という上位評価視点の更に下層に「質感」という下位評価視点が存在していることが分かる。このことから、我々は文内から抽出された下位評価視点、

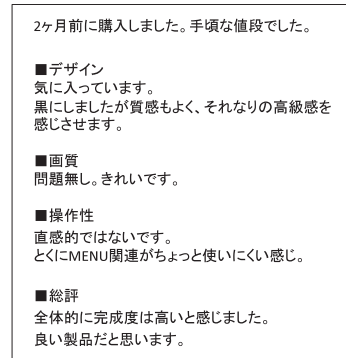


図 2 特有な書かれ方をしたレビューの例
Fig. 2 An example of review having typical format.

評価値の組に対して、更に『上位評価視点』として価格. com で指定された項目を割り当て、(上位評価視点, 下位評価視点), 評価値の組を意見として抽出する。

価格. com のレビューについて調査したところ、デジタルカメラカテゴリーの全 5774 レビューのうち、約 23% にあたる 1329 レビューがこのような特有のフォーマットで書かれていた。決して多いとはいえないが、人気商品のように多くのレビューが投稿されている場合は、このようなレビューに絞ったとしても十分な量の意見が抽出できると考えられる。しかし、レビューが少ない商品に関しては、抽出できる意見が極端に少なくなってしまう可能性があるため、通常のレビューも対象とする必要があるが、それは今後の課題とする。

また、我々の調査によると、特有のフォーマットで書かれたレビューで、かつ上位評価視点についての意見が述べられている箇所に限定した場合、同一文内に対となる下位評価視点が存在しない評価値の多くが 1 文以上前の上位評価視点に対する直接的な評価値であることが分かった。例えば、図 2 の 7 行目にある「きれい」は、2 文前に記述されている「画質」に対する評価値である。詳しい調査の結果を表 1 に示す。50 レビュー、1505 文を対象としており、667 の意見が抽出された。これを見ると、同一文内に下位評価視点が存在しない評価値の多くは、1 文以上前の上位評価視点と対となっており、1 文以上前の下位評価視点と対となるものは少ないことが分かる。そこで本論文では、『文内に下位評価視点をもたない評価値』を『1 文以上前の上位評価視点に対する評価値』とし、下位評価視点が存在しない (φ, 評価値) の組に対して上位評価

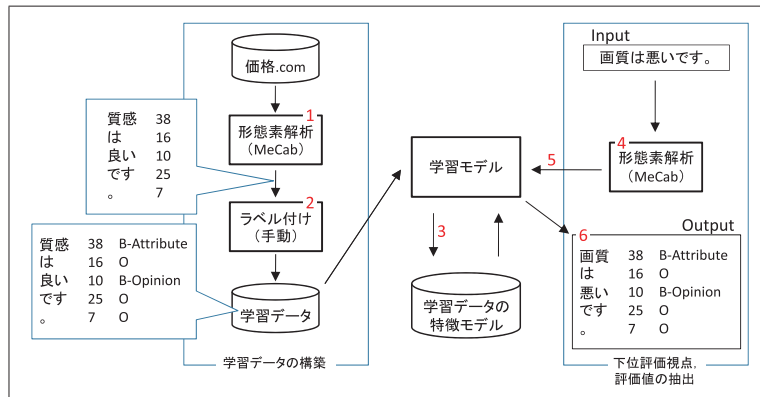


図3 下位評価視点, 評価値の抽出の流れ

Fig. 3 Flow of extraction of lower attribute and value.

表1 評価値の出現数 (対となる評価視点の位置別)

Table 1 Number of occurrences of values.

対となる評価視点の位置	出現数	割合
同一文内に下位評価視点	336	50.4%
1文以上前に『上位評価視点』	314	47.1%
1文以上前に『下位評価視点』	17	2.5%
評価視点なし (対象物への評価値)	0	0%

視点の割当を行い, \langle (上位評価視点, ϕ), 評価値 \rangle という組も意見として抽出する.

上位評価視点の同定は, 特有なフォーマット, つまり図2のようなフォーマットで書かれたレビューの構築を利用することで行う. よって本手法では, 分析対象を図2のようなフォーマットのレビューで, かつ上位評価視点についての意見が述べられている箇所に限定する. 例えば図2であれば, 最終的な意見として以下のようなものが抽出される.

- \langle (デザイン, ϕ), 気に入っている \rangle
- \langle (デザイン, 質感), よい \rangle
- \langle (デザイン, ϕ), 高級感を感じる \rangle
- \langle (画質, ϕ), 問題なし \rangle
- \langle (画質, ϕ), きれい \rangle
- \langle (操作性, ϕ), 直感的ではない \rangle
- \langle (操作性, MENU 関連), 使いにくい \rangle

「■デザイン」より上の箇所と「■総評」以降は, 上位評価視点についての意見ではないため除外している.

3.2 下位評価視点, 評価値の抽出

下位評価視点, 評価値の抽出の流れを図3に示し, 以下順番に説明する.

はじめに学習データを構築する. まず, 価格.comからレビューを一定量取得し, 全てに対して形態素解

析を行う (図3の1). 形態素解析には MeCab [7] を用いる. そして MeCab の出力に対して手でラベル付けを行い (図3の2), 学習データとする.

次に, 構築した学習データと学習モデルを用いて特徴モデルの生成を行う (図3の3).

そして最後に, 下位評価視点と評価値を抽出したい文を形態素解析し (図3の4), 学習モデルに入力として与える (図3の5) ことで, 特徴モデルにより下位評価視点と評価値が抽出される (図3の6). この抽出の結果が3.3の「下位評価視点-評価値対の同定」の入力となる.

3.2.1 本手法で用いる学習モデル

学習モデルには主に, 生成モデルに基づくものと識別モデルに基づくものの二つがあり, 生成モデルでは HMM (hidden Markov models), 識別モデルでは CRF (conditional random fields), SVM (support vector machines), MEMM (maximum entropy Markov models) 等が知られている. その中で CRF はラベル付与の問題において, HMM と MEMM に存在する問題点を解決することが知られており, 既存研究の実験においても有効性が確認されている [2]. また, SVM はノイズにの多いデータに対して頑健であるといわれており [1], ノイズの多い Web 上のレビューに対しても有効であると考えられる. そこで本研究では, CRF と SVM をラベル付与の問題に用いることとする. ただし, SVM は2値分類器であるため, そのままでは多クラスに分類することができない. そこで SVM を多クラス分類問題に適用できるように拡張された SVMmulticlass [8] を用いる.

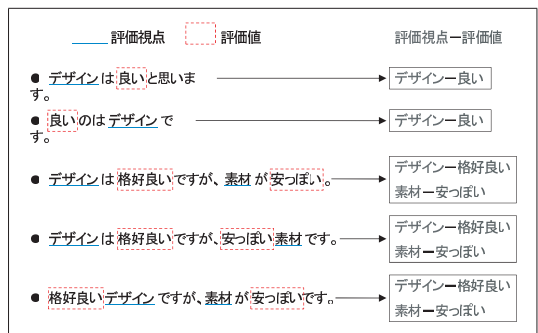


図 4 下位評価視点-評価値の例
Fig. 4 Example of identification of lower attribute.

3.3 下位評価視点-評価値対の同定

3.2 の処理により抽出された評価値に対して、同じく抽出した下位評価視点を同定する。ここでは単純に以下のルールにより同定を行う。

1. 文内の評価値が一つの場合
 - 1-1 評価値より手前に出現する下位評価視点を全て対として同定する。
 - 1-2 評価値より手前に出現しない場合、後ろに出現する下位評価視点を全て対として同定する。
2. 文内の評価値が二つ以上の場合
 - 他の評価値よりも近い距離にある下位評価視点を同定する。距離は形態素で計る。

下位評価視点と対になることができる評価値は一つのみである。同じ距離の場合、後方の評価値に同定する。同定の例を図 4 に示す。また、評価値は二つ以上の下位評価視点と対となることができる。

4. 評価実験

4.1 データセット

評価実験を行うにあたり、価格.com から複数の商品のレビューを抽出し、学習データとテストデータをそれぞれ異なる商品のレビューから構築した。以下各々について述べる。

4.1.1 学習データ

特徴モデルを生成するための学習データについて説明する。学習データは、抽出してきたレビューを形態素解析し表層文字列と品詞番号を出力したものに対して、人手でラベルを付与したものである。ラベル付与は、意見性を満たす下位評価視点-評価値対に対して付与者の主観で行った。意見性の有無について、例文を表 2 に示す。意見性とは、下位評価視点-評価値対

表 2 意見性の有無
Table 2 Example of opinion or not.

意見性	例文	形式
あり	この商品は画質が 良い。	主観的な意見
なし	画質が 良い と友人が言っていた。	伝聞
なし	この商品に比べて商品 B は 良い。	対象物違い
なし	画質が 良ければ 最高 なんだけど。	条件・仮定

表層文字列	品詞番号	ラベル
派手	40	B-Opinion
すぎ	58	I-Opinion
です	25	O
し	18	O
、	9	O
質感	38	B-Attribute
も	16	O
良く	10	B-Opinion
ない	25	I-Opinion
です	25	O
。	7	O

図 5 学習データの例
Fig. 5 An example of training data.

が意見となるか否かの指標であり、飯田ら [11] は、「記事の書き手の主観的な意見」のみが意見性をもつとしている。本論文においてもこの基準を用いる。なお、異なる対象物に対する意見も意見性はないものとして扱う。

本手法において学習データに付与するラベルは、下位評価視点を示す「Attribute」と評価値を示す「Opinion」の二つである。ただし、『B (Begin)』は表現の始まり、『I (Inside)』は表現の途中、『O (Outside)』は表現以外を表す。学習データの例を図 5 に示す。

以上の手順で「デジタルカメラ」カテゴリーの 11 レビューから抽出した 315 文を用いて学習データを作成した。

4.1.2 テストデータ

学習データの作成に用いた商品とは別の商品に対するレビューからテストデータを作成した。価格.com における、「デジタルカメラ」、「MP3 プレーヤー」、「ノートパソコン」、「エアコン」、「自動車」の各カテゴリーからレビューを抽出した。それらを形態素解析し、表層文字列と品詞番号を出力したものをテストデータとした。また、精度算出のために手動で意見を抽出し、正解データセットを構築した。テストデータの詳細を表 3 に示す。

表 3 テストデータの内訳
Table 3 The summary of test data.

ジャンル	レビュー数	文数	意見数 (正解データ数)
デジタルカメラ	86	2396	1027
MP3 プレーヤー	8	172	106
ノートパソコン	7	192	103
エアコン	15	265	103
自動車	16	304	109

4.2 実験方法

以下の三つについて精度を算出した。

1. 下位評価視点, 評価値の抽出の精度算出
 2. 意見抽出の精度算出
 3. 他ジャンルに適用した場合の意見抽出の精度算出
- まず, 4.1.1 で述べた学習データを用いて特徴モデルを生成する。その際に用いた学習モデルは, 3.2 でも述べたように「CRF」と「SVM」の二つである。そして 4.1.2 のテストデータを入力として与え, 自動的にラベルが付与されたものを正解データと照合し, 精度を算出する。

4.2.1 下位評価視点, 評価値の抽出の精度算出

教師あり学習がどの程度効果を発揮するのかを評価するため, 下位評価視点, 評価値の抽出単独での精度を算出する。

正解データと特徴モデルの出力を照合し完全に表現が一致した場合にのみ正解とした。例えば、『マクロ機能』のように『マクロ』と『機能』の二つの形態素で構成される下位評価視点は、『マクロ』を『B-Attribute』, 『機能』を『I-Attribute』と分類していなければ不正解であり, 同様に『良くない』のように『良く』と『ない』で構成される評価値は、『良く』を『B-Opinion』, 『ない』を『I-Opinion』と分類していなければ不正解である。

4.2.2 意見抽出の精度算出

抽出した下位評価視点と評価値から, パターンを用いて下位評価視点-評価値対を同定し, 最後に上位評価視点を割り当てる。具体的には, 3.3 で述べたパターンに応じて『Attribute』と『Opinion』の対を同定する。対をもたない『Opinion』も存在する。そして最後に, 抽出元のレビューの構造を認識して上位評価視点を割り当て, <(上位評価視点, 下位評価視点), 評価値> 若しくは, <(上位評価視点, ϕ), 評価値> の形式で出力する。以上の手順で抽出した意見を人手で確認し, 精度を算出する。また, 比較対象として以下の手

法の精度算出を行った。

1. BaseLine

Web 上で公開されている「日本語評価極性辞書」[12]と我々が手動で構築した下位評価視点辞書を用いた手法である。下位評価視点辞書は, テストデータとは別のレビュー 3197 文から名詞と名詞の連続を抽出し主観で登録した。辞書は 146 表現から構成される。意見の抽出方法は, 各辞書で下位評価視点, 評価値の候補を抽出し, それらが係り受け関係にあった場合に抽出する。

2. 高野ら [15] の手法

大規模テキストを用いて意見表現を収集する手法であり, 小規模な種辞書を与えることで, ブートストラップ的に評価値, または下位評価視点と評価値の組合せが抽出される。この手法の目的は表現を収集することであるが, 意見抽出の中に直接組み込むことも検討されていたため, 比較対象とした。文献の実験環境に合わせ, 15000 文から 2 サイクルの抽出処理を行った。抽出された 86 表現を用いて, 意見抽出を行う。

4.2.3 他ジャンルに適用した場合の意見抽出の精度算出

他ジャンルへ適用することで, コスト削減の可能性を評価する。本実験では, 「デジタルカメラ」カテゴリのレビューから構築した学習データを用いて, 他ジャンルのレビューから意見抽出を行い, その精度を算出する。対象ジャンルは, 「MP3 プレーヤー」, 「ノートパソコン」, 「エアコン」, 「自動車」の四つである。

4.3 実験結果

各実験の適合率と再現率を算出した。下位評価視点, 評価値の抽出単独での精度の算出は以下の式を用いた。

● 下位評価視点

$$\text{適合率} = \frac{\text{正しく抽出できた下位評価視点の数}}{\text{下位評価視点と判定した表現の数}}$$

$$\text{再現率} = \frac{\text{正しく抽出できた下位評価視点の数}}{\text{正解データの下位評価視点の数}}$$

● 評価値

$$\text{適合率} = \frac{\text{正しく抽出できた評価値の数}}{\text{評価値と判定した表現の数}}$$

$$\text{再現率} = \frac{\text{正しく抽出できた評価値の数}}{\text{正解データの評価値の数}}$$

算出した精度を表 4 に示す。

意見抽出と他ジャンルに適用した場合の精度算出は

表 4 下位評価視点、評価値の抽出の精度

Table 4 Comparison of our approach with baseline and Takano approach.

	学習モデル	適合率	再現率	F 値
下位評価視点	CRF	29.9% (203/680)	41.8% (203/486)	34.8%
	SVM	34.8% (213/612)	43.8% (213/486)	38.3%
評価値	CRF	52.5% (716/1365)	69.7% (716/1027)	59.9%
	SVM	54.6% (662/1213)	64.5% (662/1027)	59.1%

表 5 意見抽出の精度

Table 5 Precision of opinion extractions.

手法		適合率	再現率	F 値
本手法	CRF	39.6% (541/1365)	52.7% (541/1027)	45.2%
	SVM	43.4% (527/1213)	51.3% (527/1027)	47.0%
BaseLine		20.7% (133/641)	12.9% (133/1027)	15.9%
高野らの手法		17.7% (61/345)	5.9% (61/1027)	8.9%

以下の式を用いた。

$$\text{適合率} = \frac{\text{正しく抽出できた意見の数}}{\text{意見として抽出された数}}$$

$$\text{再現率} = \frac{\text{正しく抽出できた意見の数}}{\text{正解データの評価値の数 (= 意見の数)}}$$

算出した意見抽出の精度を表 5 に、他ジャンルに適用した場合の精度を表 6 に示す。

4.4 考 察

4.4.1 下位評価視点と評価値の抽出

評価値の抽出の精度に関しては、適合率が最大で約 55%、再現率が最大で約 70%となるなど、一定の成果が見られた。しかしながら、下位評価視点の抽出の精度は、適合率が最大で約 35%、再現率が最大で約 44%と、あまり良い値が得られなかった。また、どちらも再現率に比べ、適合率が低い値となった。下位評価視点と評価値の誤抽出を CRF の出力と SVM の出力からランダムに各 100 件抽出し分析した結果、共に最も多い誤抽出が、「意見性がない表現」であった。これは、「下位評価視点、評価値になり得るが、当該部分では抽出すべきではない表現」を指す。例えば、SONY 製のカメラへのレビューで、『パナソニックの DMC-FX150の方が高級感はあるなあ』という文では、一見「高級感-ある」という下位評価視点と評価

表 6 他ジャンルに適用した場合の精度

Table 6 Precision in case of the other genre.

ジャンル	学習モデル	適合率	再現率	F 値
MP3 プレーヤー	CRF	48.8% (60/123)	56.6% (60/106)	52.4%
	SVM	48.1% (52/108)	49.1% (52/106)	48.6%
ノートパソコン	CRF	42.6% (49/115)	47.6% (49/103)	45.0%
	SVM	45.1% (41/91)	39.8% (41/103)	42.3%
エアコン	CRF	35.9% (51/142)	49.5% (51/103)	41.6%
	SVM	35.9% (47/131)	45.6% (47/103)	40.2%
自動車	CRF	33.3% (52/156)	47.7% (52/109)	39.2%
	SVM	38.1% (51/134)	46.7% (51/109)	42.0%

値が存在しているように見えるが、パナソニックの「DMC-FX150」という別の対象物への意見であり、意見性がないと判断されなければならない。しかし、このような誤抽出が本手法では目立った。また、同様に多く見られたのが、『良くない』のように複数の形態素から構成される表現の一部しか抽出できず、結果として誤抽出となってしまう例である。以上のような誤りを防ぎ、精度を向上させるためには、学習データの追加や、文献[5]のような意見性判定の処理の導入が有効であると考ええる。

4.4.2 意見抽出と他ジャンルに適用した場合の意見抽出

意見抽出については、BaseLine、高野らの手法よりも精度が良く、有用性が確認された。精度が高まった大きな要因として、様々な表現に対応できる点が挙げられる。BaseLine、高野らの手法は、共に辞書を用いているが、それらの辞書に登録されていない表現の出現が目立った。その一例として、『EXRAUTO モード』や『スイングパノラマ』といったような商品独自の機能名がある。このような機能名は、新商品が出るたびに新しく生成される可能性があるため、あらかじめ辞書に登録しておくことは困難である。本手法では、そのような機能名が学習データに存在しない状態でも抽出できていた。表 6 に示した他ジャンルでの抽出の精度を見ても、デジタルカメラカテゴリーに適用した場合とほぼ変わらない精度で抽出できていることから、本手法は表現の異なる他ジャンルにも適用できる可能性があると考えられる。

4.4.3 下位評価視点-評価値対の同定

表5で示したように、比較手法に比べ良い精度が出ているが、実際にシステムに应用するためにはまだ十分とはいえない。精度を落とした要因は、下位評価視点と評価値の抽出の精度が低かったことの他に、対の同定方法にあると考える。今回は対の同定方法として、下位評価視点を、最も距離に近い評価値に対して同定する方法を用いたが、それでは正しく同定できない例が多く見られた。例えば、『ノイズが少なく、画質には非常に満足しています』という文の場合、本来は「ノイズ-少ない」、「画質-満足」という意見が抽出されなければならないが、「ノイズ」と「画質」がともに「少ない」と対になってしまうため、結果的に二つの誤抽出が発生することになり、適合率、再現率共に下げる結果となってしまう。そのため、距離だけでなく他の要素を取り入れた同定方法を検討する必要がある。飯田ら[11]のトーナメントモデルは、表層文字列、品詞、距離、かかわり受け関係にあるか否かを素性とした機械学習により対同定を行う手法であり、有効であると考えられるが、評価値が一つの下位評価視点としか対となれないなどの問題点があるため、手法の改良が必要である。

4.4.4 CRF と SVM について

実験結果を見ると、下位評価視点の抽出ではSVMが、評価値の抽出ではCRFの精度がそれぞれ高くなっている。また意見抽出ではSVMだが、他ジャンルに適用した場合ではCRFの方が精度が良い結果となっている。実験結果からは一概にどちらが良いのかはいえないが、基本的にCRFは出力数が多く、それに伴って再現率が高くなっていることが分かる。反対にSVMはCRFに比べて出力数が少なく、適合率が高くなっている結果が多い。これは、SVMがノイズに強く、ノイズに対してのラベルを付与し難いことに起因していると考えられる。ノイズが多く入っていると考えられる長文においては、SVMがCRFに比べ正しいラベルを付与している印象にある。『フィルムを使っておられた方でしたら、ISO800や1600のフィルムを使ったときのように、ややざらついてはいるが、粒子が整っているので画の雰囲気壊さない、と説明すれば御理解頂けるでしょうか』という長文に対してのCRFとSVMのラベル付与の一部を図6に示す。長い文の途中にある意見に対して、SVMは正しいラベルを付与できていることが分かる。

CRF			SVM		
やや	34	0	やや	34	0
ざらつい	31	0	ざらつい	31	0
て	18	0	て	18	0
は	16	0	は	16	0
いる	33	0	いる	33	0
が	18	0	が	18	0
、	9	0	、	9	0
粒子	38	0	粒子	38	B-Attribute
が	13	0	が	13	0
整っ	31	0	整っ	31	B-Opinion
て	18	0	て	18	I-Opinion
いる	33	0	いる	33	I-Opinion
ので	18	0	ので	18	0

図6 長文におけるラベル付与の例

Fig.6 A labeling result of along sentence.

4.4.5 比較手法に関して

高野らの手法がBaseLineを下回った一番の要因は、辞書の表現の少なさである。しきい値の設定の関係により文献上の実験環境に合わせたが、本来はより大規模なテキストが必要であると考えられる。また、ほぼ自動抽出であるため、意見とはならない表現の抽出も多く、今回抽出した86表現の適合率は64%であった。この点も精度低下の原因である。システムに直接組み込む場合、更なるフィルタリング処理が必要である。

4.4.6 本手法のコスト

本手法における学習データの構築のコストを、辞書を用いた手法と比較する。飯田ら[11]の手法は、あらかじめ辞書に用例を登録し、それらとマッチしたものを意見として抽出している。辞書には約2000~3000の用例が登録されている。それに対して本手法では、約300の学習データで、飯田らの手法と同程度の精度を出すことが可能である。また飯田らの手法では、実験で用いた「デジタルカメラ」と「車（自動車）」、各々のジャンルに対して個別の辞書の構築を必要としている。それに対して本手法では、ジャンルごとに学習データを構築せず、一組の学習データで複数のジャンルに適用することが可能である。以上のことから、本手法は辞書を用いた手法と比較して、低いコストで意見の抽出が可能である。

4.4.7 上位評価視点導入の効果

本手法は、評判分析システムに应用することを目的とし、抽出した意見に対して、グルーピングするための「上位評価視点」を付与している。そのために、価格.comの特有なフォーマットで書かれたレビューを利用したが、このようなレビューの場合、1文以上前の上位評価視点に対する意見が目立ち、結果的に精度向上につながった。例えば、3.1で示した図2のよう

なレビューの場合、〈(画質, ϕ), 問題なし〉と〈(画質, ϕ), きれい〉という形で抽出されることによって、「問題なし」、「きれい」という評価値が「画質」という評価視点に対するものであることが分かるが、上位評価視点を付与しなければ、「問題なし」と「きれい」が何に対する評価値なのか分からず、正しい意見抽出とならない。我々がランダムに抽出した 1027 の意見のうち、約 60%にあたる 621 の意見が、このように上位評価視点と評価値のみから構成される意見であった。

4.4.8 本手法の適用可能範囲

本手法を適用できるジャンルの範囲を確認するために、実験で用いたジャンルの共通性について考察する。

まず、学習データとして用いた「デジタルカメラ」の上位評価視点と、他ジャンルとして取り上げた「MP3 プレーヤー」、「ノート PC」、「エアコン」、「自動車」の各々の上位評価視点がどの程度共通しているかを調べた。「MP3 プレーヤー」は 7 項目中 3 項目、「ノート PC」は 8 項目中 4 項目、「エアコン」は 6 項目中 1 項目、「自動車」は 7 項目中 0 項目が「デジタルカメラ」の上位評価視点と共通の項目であった。「MP3 プレーヤー」、「ノート PC」は約半数が共通項目であることから、「デジタルカメラ」と近いジャンルであると考えられるが、「エアコン」、「自動車」は共通項目がほとんどなく、共通性は低いと考えられる。また、レビューの文中の表現においても、例えばエアコンであれば「室外機」や「ナノイー」、「カビ臭」といった他のジャンルではほとんど用いられない特有の表現が意見の一部として用いられており、本手法ではこれらの抽出もできていた。このように、既存手法では個別の辞書を必要とするような状況であっても、本手法は一つの学習データで、更なる学習データを用いずに同程度の精度を示せたことから、共通性が低いジャンルに対しても適用できる可能性がある。

5. 関連研究

以下、二つの観点から関連研究を述べる。

5.1 辞書をあらかじめ用意する意見抽出

辞書を用いた意見抽出の研究として、立石ら [10]、飯田ら [11]、Kobayashi ら [5] の研究がある。これらの研究は、あらかじめ構築した辞書を用いて評価視点、評価値を抽出しているが、ジャンルごとに必要となる辞書を構築することは多大なコストとなる。それに対して本研究では、一つのジャンルから構築した学習データで複数の他ジャンルの製品レビューから意見

を抽出できることから、低コストで多くのジャンルからの意見抽出を実現できる可能性があると考えられる。以下、上記の関連研究を四つのタスクに分けて説明する。

(1) 評価視点候補、評価値候補を同定する

対となる評価視点と評価値の候補を網羅的に抽出するタスクである。このタスクで同定された評価視点候補と評価値候補が対となる。評価視点候補、評価値候補のいずれにおいても、あらかじめ辞書を構築しておく、当てはまる表現を全て候補として抽出している。2.3 で述べたように、無数にあるジャンル全てに対して辞書を構築することは現実的ではない。

(2) 評価値候補が評価視点をもつか否かを判定する (対判定)

(1) において同定された評価値候補が、対となる評価視点をもつか否かを判定するタスクである。Kobayashi らの研究では機械学習を用いて判定を行い、立石ら、飯田らの研究では (3) のタスクにおいて評価視点が同定されなかった場合に対をもたないと判定している。

(3) 評価値候補に対して、対となる評価視点候補を同定する (評価視点-評価値対同定)

(1) において同定された評価視点候補と評価値候補の対を同定する。立石らは、抽出のルールをあらかじめ設定し、マッチしたペアを対として抽出している。飯田らはこのタスクに対してトーナメントモデルを提案している。この手法は、機械学習を用い、一つの評価値に対して複数の評価視点候補の中から最優秀の候補を決定する手法である。Kobayashi らも評価視点同定にはトーナメントモデルを用いている。

全ての手法は評価視点-評価値対を 1 対 1 に限定している。しかし実際には二つ以上の評価視点をもつ評価値も存在する。

(4) 評価視点-評価値対が意見性をもつか否かを判定する

最終的に同定した評価視点と評価値の対が意見となるか否かの判定を行うタスクである。飯田らは、「記事の書き手の主観的な意見」のみが意見性をもつと設定しており、機械学習を用いて判定している。

5.2 辞書をあらかじめ用意しない意見抽出

高野ら [15] の手法は、あらかじめ大規模な辞書を用意せず、数単語の種辞書から辞書を自動的に構築する。しかし誤抽出も多く、システムに直接組み込むためにはフィルタリングを更に洗練する必要がある。表 5 で示したように、高野らの手法と比較して、本手法の精

度はF値で約38%程度向上している。

また、辞書を用いずに評価視点を抽出する手法として、Huら[3]の手法がある。Huらはレビュー文書に対して、相関ルールマイニングの手法を用いて、評価文書集合中に頻出し、かつ評価表現から近い位置に出現する名詞あるいは名詞句を評価視点として抽出している。しかし、評価視点は頻度が低いものが多数存在するため、抽出できないものが出ると考えられる。例えば、今回我々がランダムに手動で抽出した1027の意見の中には、255種の下位評価視点が存在し、そのうちの188種が1度しか出現しないものであった。本手法では、頻度に依存せずに評価視点を抽出することが可能である。

6. む す び

本論文では、レビューページを対象とし、教師あり学習を用いた意見抽出手法の提案と、特有の書かれ方をしているレビューの構造を用いて『上位評価視点』を割り当ててことを行った。BaseLineの手法と既存手法を対象として比較を行ったところ、精度の観点から本手法の有効性が確認された。しかし、システムに応用するためには十分な精度とはいえないため、今後、精度の向上のために、4.4で述べた学習データの追加や意見性判定の処理の導入を検討する。また、下位評価視点-評価値対の同定についても、距離のみではなく、他の要素を組み合わせることで対を同定する必要がある。今後そのような手法を検討していく。また、本論文では対象レビューサイトを「価格.com」としたが、下位評価視点、評価値の抽出と下位評価視点-評価値対の同定は、商品ごとにレビュースペースが設けられていれば、どのようなレビューサイトでも対応可能であると考えられる。しかし、上位評価視点の同定は、今回のようにあらかじめ評価項目が設定されており、かつ特有のフォーマットで記述されている必要がある。今後は、評価項目が設定されていないレビューサイトにも適用可能な手法を検討していく。

謝辞 本研究を遂行するにあたり、研究の機会と議論・研鑽の場を提供して頂き、御指導頂いた国立情報学研究所/東京大学本位田真一教授をはじめ、活発な議論と貴重な御意見を頂いた研究グループの皆様に感謝致します。

文 献

- [1] B. Scholkopf, K.K. Sung, C.J.C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing sup-

port vector machines with Gaussian kernels to radial basis function classifiers," IEEE Trans. Signal Process., vol.45, no.11, pp.2758-2765, Nov. 1997.

- [2] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to Japanese morphological analysis," Proc. EMNLP2004, pp.230-237, Barcelona, Spain, July 2004.
- [3] M. Hu and B. Lie, "Mining and summarizing customer reviews," Proc. Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.168-177, Seattle, USA, Aug. 2004.
- [4] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the Web," Proc. 14th International World Wide Web Conference, pp.324-351, Chiba, Japan, May 2005.
- [5] N. Kobayashi, R. Iida, K. Inui, and Y. Matsumoto, "Opinion mining as extraction of attribute-value relations," Lect. Notes Comput. Sci. (LNCS), vol.4012, pp.470-481, 2006.
- [6] T. Kudo and Y. Matsumoto, "Cabocha: Yet another Japanese dependency structure analyzer," <http://code.google.com/p/cabocha/>, Aug. 2010.
- [7] T. Kudo, "MeCab: Yet another part-of-speech and morphological analyzer," <http://mecab.sourceforge.net/>, Aug. 2010.
- [8] T. Joachims, "SVM-Light Support Vector Machine," http://svmlight.joachims.org/svm_struct.html, Nov. 2010.
- [9] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, "テキストマイニングによる評価表現の収集," 情処学研報, 自然言語処理研究会報告, vol.2003, no.23, pp.77-84, March 2004.
- [10] 立石健二, 福島俊一, 小林のぞみ, 高橋哲郎, 藤田 篤, 乾健太郎, 松本裕治, "Web 文書集合からの意見情報抽出と着眼点に基づく要約生成," 情処学研報, vol.2004, no.93, pp.1-8, Sept. 2004.
- [11] 飯田 龍, 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, "意見抽出を目的とした機械学習による属性-評価値対同定," 情処学研報, 自然言語処理研究会報告, vol.2005, no.1, pp.21-28, Jan. 2005.
- [12] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, "意見抽出のための評価表現の収集," 自然言語処理, vol.12, pp.203-222, 2005.
- [13] 小林のぞみ, 乾健太郎, 松本裕治, "意見情報の抽出/構造化のタスク仕様に関する考察," 情処学研報, 自然言語処理研究会報告, vol.2006, no.1, pp.111-118, Jan. 2006.
- [14] 乾 孝司, 奥村 学, "テキストを対象とした評価情報の分析に関する研究動向," 自然言語処理, vol.13, no.3, pp.201-241, 2006.
- [15] 高野敦子, 池奥渉太, 北村泰彦, "因果関係に着目した口コミ Web サイトからの評価表現抽出," 人工知能学会論文誌, vol.24, no.3, C, pp.322-332, 2009.

付 録

評価実験のデータセット

- 学習データ, テストデータ, 正解データ

<https://docs.google.com/leaf?id=0B2HokKkjQrfHZGE4YTYwYWQtNWU2My00MzY3LWJhNzEtZTIzYjgxZWYzMTNh&authkey=CPT4nuAK&hl=ja>

(平成 23 年 1 月 13 日受付, 5 月 21 日再受付)



大須賀昭彦 (正員)

1981 上智大・理工・数学卒。同年 (株) 東芝入社。同社研究開発センター, ソフトウェア技術センターなどに所属。1985～1989 (財) 新世代コンピュータ技術開発機構 (ICOT) 出向。2007 より, 電気通信大学大学院情報システム学研究科教授。工博 (早稲田大学)。主としてソフトウェアのためのフォーマルメソッド, エージェント技術の研究に従事。1986 年度情報処理学会論文賞受賞。情報処理学会, 日本ソフトウェア科学会, 人工知能学会, IEEE CS 各会員。



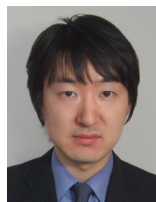
田中 俊行

2009 東京工芸大・工・コンピュータ応用学科卒。現在, 電気通信大学大学院情報システム学研究科社会知能情報学専攻修士課程在学。



グエン ミンティ

2009 電気通信大学大学院情報システム学研究科修士課程了。現在, 同大学院博士課程社会知能情報学専攻在学。主にユーザの状況に応じた行動推薦に関する研究に従事。



中川 博之 (正員)

1997 阪大・基礎工・情報工学卒。同年鹿島建設 (株) に入社。2007 東京大学大学院情報理工学系研究科修士課程了。2008 同大学院博士課程中退。同年より電気通信大学助教。現在に至る。エージェント及び自己適応システム開発手法の研究に従事。



田原 康之

1991 東京大学大学院理学系研究科数学専攻修士課程了。同年 (株) 東芝入社。1993～1996 情報処理振興事業協会に出向。1996～1997 英国 City 大学客員研究員。1997～1998 英国 Imperial College 客員研究員。2003 国立情報学研究所入所。2008 より電気通信大学准教授。博士 (情報科学) (早稲田大学)。エージェント技術, 及びソフトウェア工学などの研究に従事。情報処理学会, 日本ソフトウェア科学会会員。